Table 22 Summary of ML for flow feature-based anomaly detection

Ref.	ML Technique	Dataset	Features	Evaluation	
				Settings	Results
Kayacik et al. [232]	Unsupervised Hierarchical SOM (*Offline*)	KDD Cup [257]	6 TCP features	-494,021 training records -311,029 records in test set 1 -4,898,431 records in test set 2 -Platforms: SOM-Toolbox [12] & SOM PAK [250] -3-level SOM w/ # Epochs: 4000	DR Test-set 1: 89% FP Test-set 1: 4.6% DR Test-set 2: 99.7% FP Test-set 2: 1.7%
Kim et al. [242]	Supervised SVM (*Offline*)	KDD Cup [257]	selected using GA	Training set: kddcup.data.gz [257] Testing set: corrected.gz [257] -Detect only DoS attacks -10-fold cross validation -GA ran for 20 generations	DR w/ Neural Kernel: 99% DR w/ Radial Kernel:87% DR w/ Inverse Multi-Quadratic Kernel: 77%
Jiang et al. [220]	Unsupervised Improved NN (*Offline*)	KDD Cup [256, 257]	all 41 features	-40,459 training records -429,742 testing records -Cluster Radius Thresh r=[0.2-0.27]	DR DoS: 99.10%%99.15 DR Probe: 64.72%80.27% DR U2R: 25.49%60.78% DR R2L 6.34%8.67% DR new attacks: 32.44%42.12% FP: 0.05%1.30%
Zhang et al. [495]	Unsupervised Random Forests (*Offline*)	KDD Cup [257]	40 features labeled by service type	-4 datasets used with % of attack connections: 1%, 2%, 5%, 10% -Platform used: Weka [288]	1% attacks: FP: 1% DR: 95% 10% attacks: FP: 1% DR: 80%
Ahmed et al. [7]	Supervised Kernel Function (*Online*)	From Abilene backbone network	number of packets, number of individual IP flows	-2 timeseries binned at 5 min intervals -Timeseries dimensions = FxT -F = 121 flows, T = 2016 timesteps	T#1 DR: 21/34-30/34 FP:0-19 T#2 DR:28/44-39/44 FP:5-16
Shon et al. [411]	Unsupervised Soft-margin SVM and OCSVM (*Offline*)	KDD Cup [257] Data collected from Dalhousie U.	selected using GA	-SVM Toolkits [88, 396] -100,000 packets for training -1,000-1,500 packet for testing -GA run for 100 generations 3-cross fold validation	KDD w/ 9 attack types DR: 74.4% Dalhousie Dataset DR: 99.99% KDD w/ 9 attack types FN:31.3% Dalhousi Dataset FP:0.01%
Giacinto et al. [165]	Unsupervised Multiple Classifiers (*Offline*)	KDD Cup [257]	29 features for HTTP 34 features for FTP 16 features for ICMP 31 features for Mail 37 features for Misc 29 features for Private&Other	-494,020 training records -311,029 testing records -1.5% of data records is attacks	v-SVC DR: 67.31%-94.25% v-SVC FP: 0.91%-9.62%
Hu et al. [198]	Supervised Decision stumps with AdaBoost (*Offline*)	KDD Cup [257]	all 41 features	-494,021 training records -311,029 testing records -Pentium IV with 2.6-GHz CPU and 256-MB RAM -Platform used Matlab 7	DR: 90.04%-90.88% FP: 0.31%-1.79% Mean Training time: 73 sec
Muniyandi et al. [327]	Unsupervised K-Means, C4.5 DT (*Offline*)	KDD Cup [257]	all 41 features	-15,000 training records -2,500 testing records -Intel Pentium Core 2 Duo CPU 2.20GHz, 2.19GHz, 0.99GB of RAM w/ Microsoft Windows XP (SP2) -Platform: Weka 3.5 [288]	DR: 99.6% FP: 0.1% Precision: 95.6% Accuracy: 95.8% F-measure: 94.0%
Panda et al. [345]	Unsupervised RF, ND, END (*Offline*)	NSL-KDD [438]	all 41 features	-25,192 training instances -IBM PC of 2.66GHz CPU with 40GB HDD and 512 MB RAM -10-fold cross validation	TP: 99.5 FP: 0.1% F-measure: 99.7% Precision: 99.9% Recall 99.9% Time to build model: 18.13 sec
Boero et al. [64]	Supervised RBF-SVM (*Offline*)	Normal: from U. of Genoa Malwares: [126, 292, 348, 351]	7 SDN OpenFlow features	-RBF Complexity par: 20 -RBF kernel par: 2	Normal-TP: 86% Normal-FP: 1.6% Malware-TP: 98.4% Malware-FP: 13.8%

empirical evaluation, the resultant IDS consist of using DT for U2R, SVM for DoS, and and DT-SVM to detect normal traffic. The ensemble of the 3 methods together (with a voting mechanism) is used to detect probing and R2L attacks. The resultant accuracy for each class is presented in Table 21.

Table 23 Summary of ML for payload-based anomaly detection

Ref.	ML Technique	Dataset	Features	Evaluation	
				Settings	Results
Zanero et al. [493]	Unsupervised A two-tier SOM-based architecture (*Offline*)	Normal: KDD Cup [257] Attack: Scans from Nessus [44]	Packet headers and payload	-2,000 training packets -2,000 testing packets -10x10 SOM trained for 10,000 epochs -Platform used: SOM toolbox [12]	Improves DR by 75% over 1-tiered S.O.M
Wang et al. [459]	Unsupervised Centroid model (*Offline*)	KDD Cup [257] & CUCS	Payload of TCP traffic	-2 weeks training data -3 weeks testing data -Inside network TCP data only -Incremental learning	DR w/ payload of a packet: 58.8% DR w/ first 100 bytes of a packet: 56.7% DR w/ last 100 bytes of a packet: 47.4% DR w/ all payloads of a con: 56.7% DR w/ first 1000 bytes of a Con: 52.6% Training time: 4.6-26.2 sec Testing time: 1.6-16.1 sec
Perdisci et al. [356]	Supervised Ensemble of single-class SVM (*Offline*)	Normal: KDD Cup [257] Normal: GATECH Attack: CLET [117] Attack: PBA [149] Generic [204]	Payload	-50% of dataset for training -50% of dataset for testing -11 OCSVM trained with 2_v-grams; v=1...10 -5-fold cross validation on KDD cup -7-fold cross validation on GATECH -2 GHz Dual Core AMD Opteron Processor and 8GB RAM	Generic DR w/ FP $10^{-}5$: 60% shell-code DR w/ FP $10^{-}5$: 90% CLET DR w/ FP $10^{-}5$: 90% Detection time KDD Cup: 10.92 ms Detection time GATECH: 17.11 ms
Gornitz et al. [171]	Supervised SVDD (*Online*)	Normal: from Fraunhofer Inst. Attack: Metasploit	payload	-2,500 training network events -1,250 testing network events -Active Learning -Fraction of Labeled data: 1.5%	DR: 96% FP: 0.0015%

Stein et al. [421] employ DT with GA. The goal of GA is to pick the best feature set out of the 41 features provided in KDD'99 dataset. DT with GA is performed for every category of attacks, rendering a total of 4 DTs. The average error rate achieved by each DT at the end of 20 runs is reported in Table 21. Another interesting ensemble learning approach is the one proposed in [90], where the ensemble is composed of pairs of feature set and classification technique. More specifically, BN and CART classification techniques are evaluated on the KDD'99 dataset with different feature sets. Markov blanket [353] and Gini [76] are adopted as feature selection techniques for BN and CART, respectively. Markov blanket identifies the only knowledge needed to predict the behavior of a particular node; a node here refers to the different categories of attacks. Gini coefficient measures how well the splitting rules in CART separates between the different categories of attacks. This is achieved by pruning away branches with high classification error. For BN, 17 features out of 41 are chosen during the data reduction phase. For CART, 12 variables are selected. CART and BN are trained on the 12 and 17 features set, as well as 19 features set from [326]. They describe the final ensemble method using pairs (#features, classification), which delineates the reduced feature set and the classification technique that exhibits the highest accuracy for the different categories of attacks and normal traffic. The ensemble model achieves 100% accuracy for normal (12 features set, CART), probe (17 features set, CART), and DoS (17 features set, Ensemble), and 84% accuracy for U2R (19 features set, CART), and 99.47% accuracy for R2L (12 features set, Ensemble).

Miller et al. [314] also devise an ensemble method but based on NB classifiers, denoted as Multi-perspective Machine Learning (MPML). The key idea behind MPML is that an attack can be detected by looking at different network characteristics or "perspective". These characteristics in turn are represented by a subset of network features. Hence, they group the features of a perspective together, and train a classifier using each feature set. The intuition behind this approach is to consider a diverse and rich set of network characteristics (each represented by a classifier), to enhance the overall prediction accuracy. The predictions made by each classifier are then fed to another NB model to reach a consensus.

A limitation of the aforementioned approaches is that they are all employed offline, which inhibits their application in real life. A few related works focused on the training and detection times of their IDS. Most classifiers (e.g., image, text recognition systems) require re-training from time to time. However, for IDSs this retraining may

Table 24 Summary of deep and reinforcement learning for intrusion detection

Ref.	ML Technique	Dataset	Features	Evaluation	
				Settings	Results
Cannady et al. [85]	RL CMAC-NN (*Online*)	Prototype Application	Patterns of Ping Flood and UDP Packet Storm attacks	-3 Layers NN -Prototype developed w/ C & Matlab	Learning Error: 2.199-1.94^{-07}% New Attack Error:2.199-8.53^{-14}% Recollection Error: 0.038-3.28^{-05}% Error after Refinement: 1.24%
Servin et al. [407]	RL Q-Learning (*Online*)	Generated using NS-2	Congestion, Delay, and Flow-based	-Number of Agents: 7 -DDoS attacks only -Boltzmann's rules for E2	FP: 0-10% Accuracy:∼ 70%-∼ 99% Recall: ∼ 30%-∼ 99%
Li et al. [273]	DL DBN w/ Auto-Encoder (*Offline*)	KDD Cup [257]	all 41 features	-494,021 training records -311,029 testing records -Intel Core Duo CPU 2.10 GHz and 2GB RAM -Platform used: Matlab v.7.11 -3 Layers Encoder: 41,300,150,75,*	TPR: 92.20%-96.79% FPR: 1.58%-15.79% Accuracy: 88.95%-92.10% Training time: [1.147-2.625] sec
Alom et al. [14]	DL DBN (*Offline*)	NSL-KDD [438]	39 features	-25,000 training & testing records	DR w/ 40% data for training: 97.45% Training time w/ 40% data for training: 0.32 sec
Tang et al. [436]	DL DNN (*Offline*)	NSL-KDD [438]	6 basic features	-125,975 training records -22,554 testing records -3-Layers DNN: 6,12,6,3,2 -Batch Size: 10 # Epochs: 100 -Best Learning Rate: 0.001	Accuracy: 72.0%5-75.75% Precision: 79%-83% Recall: 72%-76% F-measure: 72%-75%
Kim et al. [245]	DL LSTM-RNN (*Offline*)	KDD Cup [257]	all 41 features	-1,930 training data records -10 test datasets of 5000 records -Intel Core I7 3.60 GHZ, RAM 8GB, OS Ubuntu 14.04 -# Nodes in Input Layer: 41 -# Nodes in Output Layer: 5 -Batch Size:50 #Epoch:500 -Best Learning Rate:0.01	DR: 98.88% FP: 10.04% Accuracy: 96.93%
Javaid et al. [213]	DL Self-taught Learning (*Offline*)	NSL-KDD [438]	all 41 features	-125,973 training records -22,544 testing records -10-fold cross validation	2-class TP: 88.39% 2-class Precision: 85.44% 2-class Recall: 95.95% 2-class F-measure: 90.4%

Table 25 Summary of ML for Hybrid Intrusion Detection

Ref.	ML Technique	Dataset	Features	Evaluation	
				Settings	Results
Mukkamala et al. [325]	Supervised RBF-SVM (*Online*)	KDD cup [257]	all 41 features	7,312 training records -6,980 testing records -Platform used: SVMLight [224]	Accuracy: 99.5% Training time: 17.77 sec Testing Time: 1.63 sec
Zhang et al. [494]	Hybrid Hierarchical-RBF (*Online*)	KDD Cup	all 41 features	-32,000 training records -32,000 testing records	SHIDS Normal DR:=99.5% SHIDS Normal FP: 1.2% SHIDS Attack DR: [98.2%-99.3%] SHIDS Attack FP: [0%-5.4%] PHIDS level 1 DR: 99.8% PHIDS level 1 DR:1.2% PHIDS level 2 DR:[98.8%-99.7%] PHIDS level 2 FP:[0%-4%] PHIDS level 3 DR: 86.9% PHIDS level 3 FP: 0% Training time: 5 min
Depren et al. [116]	Hybrid SOM w./ J.48 (*Offline*)	KDD Cup	6 basic features for SOM all 41 features for J.48	-10-fold cross validation -Two-phases SOM Training -Phase 1 learning rate:0.6 -Phase 2 learning rate: 0.05 -Confidence Val. for J.48 pruning: 25%	DR: 99.9% Missed Rate: 0.1% FP: 1.25%